

i IN3050 - Introduction to Artificial Intelligence and Machine Learning

Spring 2025

May 27

03:00 PM - 07:00 PM (4 hours)

All questions should be answered!

Each question is assigned a weight which is indicated.

The maximum number of points for the whole set is 100 points.

The course teachers will visit the exam room once around the middle of the exam time.

Multiple Choice questions make up 70% of the points on the exam. For these, exactly one alternative is correct, unless otherwise specified.

Permitted materials: None

An on-screen calculator is available.

You may answer in English, Norwegian, Danish or Swedish.

1 In Search/Optimization, what is **true** about exploration?

Select one alternative:

- ☒ Exhaustive search is a purely exploratory technique
- ☐ Exploration is only used for continuous optimization
- ☐ Gradient Descent does more exploration than exploitation
- ☐ Hill climbing is a purely exploratory technique



Maximum marks: 2

2 Which is **true** about optimization algorithms?

Select one alternative:

- ☒ Simulated annealing balances exploitation and exploration with the temperature parameter ✓
- ☐ Gradient descent can be used for discrete optimization
- ☐ Hillclimbing has more randomness than simulated annealing
- ☐ Simulated annealing only selects neighbor solutions that are better than the current solution

Maximum marks: 2

3 Which is typically **not** a part of an evolutionary algorithm (EA)?

Select one alternative:

- ☐ A termination criterion
- ☐ A fitness function
- ☒ A temperature control ✓
- ☐ A population

Maximum marks: 2

4 Which is **not** a variation operator in Evolutionary Algorithms?

Select one alternative:

- ☐ Mutation
- ☐ Recombination
- ☐ Crossover
- ☐ Fitness Sharing



Maximum marks: 2

5 What property of the elements of a solution do we often try to conserve when doing mutation and crossover on a permutation-based genotype?

Select one alternative:

- ☐ Concurrency
- ☐ Gradients
- ☐ Sum
- ☐ Adjacency



Maximum marks: 2

6 Which is **true** about Selection in Evolutionary Algorithms?

Select one alternative:

- ☒ Stochastic universal sampling is a form of fitness proportionate selection ✓
- ☐ In tournament selection, larger tournament sizes result in a lower selection pressure
- ☐ Evolutionary algorithms always select the individual with the highest fitness
- ☐ Evolutionary algorithms with parent selection do not need survivor selection

Maximum marks: 2

7 Assume we have a population size of 5 in an EA, and are about to select individuals for the next generation. The individuals we are selecting among are [A: fitness 3, B: fitness 6, C: fitness 9, D: fitness 12, E: fitness 15]. How many copies of individual C can we expect (on average) in the next generation, assuming we use Fitness Proportionate Selection?

Select one alternative:

- ☐ 2
- ☐ 5
- ☒ 1 ✓
- ☐ 3

Maximum marks: 2

8 Which is **not** a method to encourage diversity in evolutionary algorithms?

Select one alternative:

☐ Fitness Sharing

☐ Elitism



☐ Crowding

☐ Speciation

Maximum marks: 2

9 Which **two** differ most between multiobjective and single-objective evolutionary algorithms?

Select exactly two alternatives:

☐ Variation operators

☐ Selection Methods



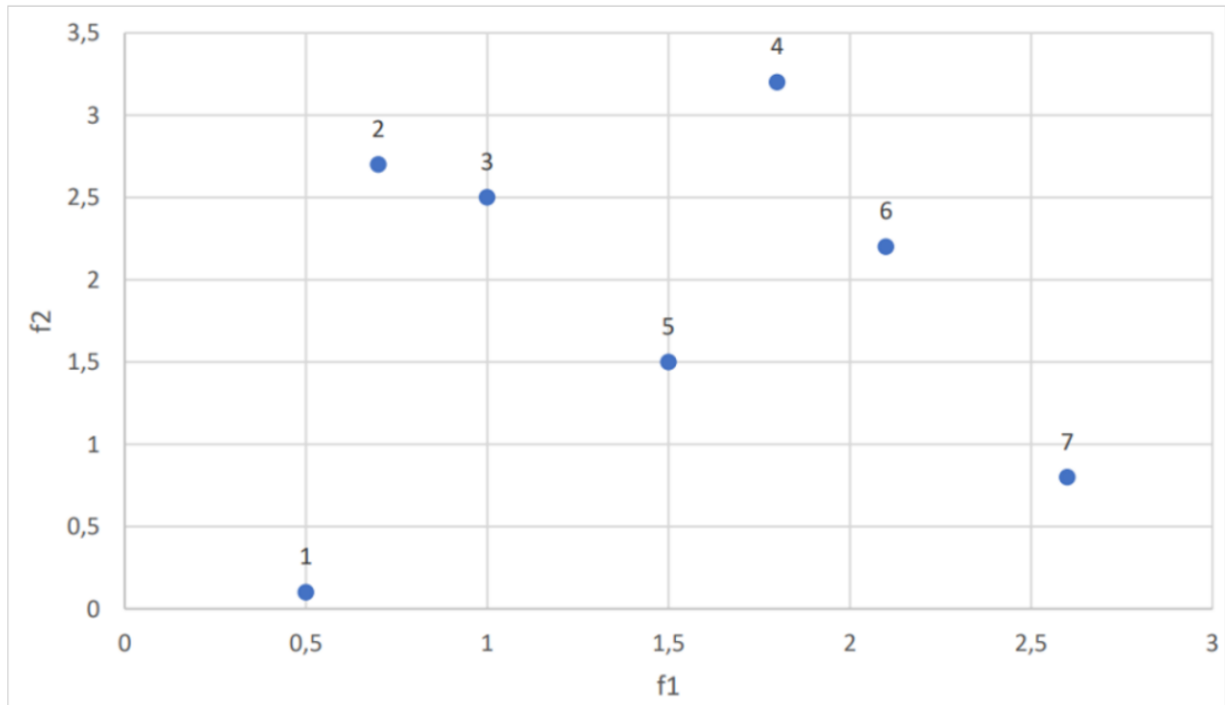
☐ Techniques for diversity maintenance



☐ Genetic representations

Maximum marks: 2

10



Assume we have a multiobjective problem with 2 objectives, f_1 and f_2 . We are aiming to maximize f_1 , and minimize f_2 . Below 7 solutions to this problem are plotted. Which **two** solutions are pareto optimal?

Select exactly two alternatives:

☐ 5

☐ 1



☐ 4

☐ 6

☐ 2

☐ 7



☐ 3

Maximum marks: 2

- 11** Consider a Reinforcement Learning environment where we are trying to learn Q-values for state/action pairs. The environment has 2 states (S1 and S2), and 2 possible actions in each state (A1 and A2). After exploring the environment, we have learned the following Q-table:

State/action	Q-value
S1/A1	0.2
S1/A2	0.6
S2/A1	0.0
S2/A2	0.8

Assume we are in state S1 and perform action A2. We observe a reward of 1 and notice that we next find ourselves in state S2. Calculate the updated Q-value $Q(S1, A2)$ assuming a discount factor of 0.9 and a learning rate of 0.1. Assume we are using the normal Q-learning update function. (answer with 2 decimal numbers)

Select one alternative:

- ☐ 0.91
- ☐ 0.51
- ☐ 0.31
- ☐ 0.71



Maximum marks: 2

- 12** Which is **not** true about SARSA?

Select one alternative:

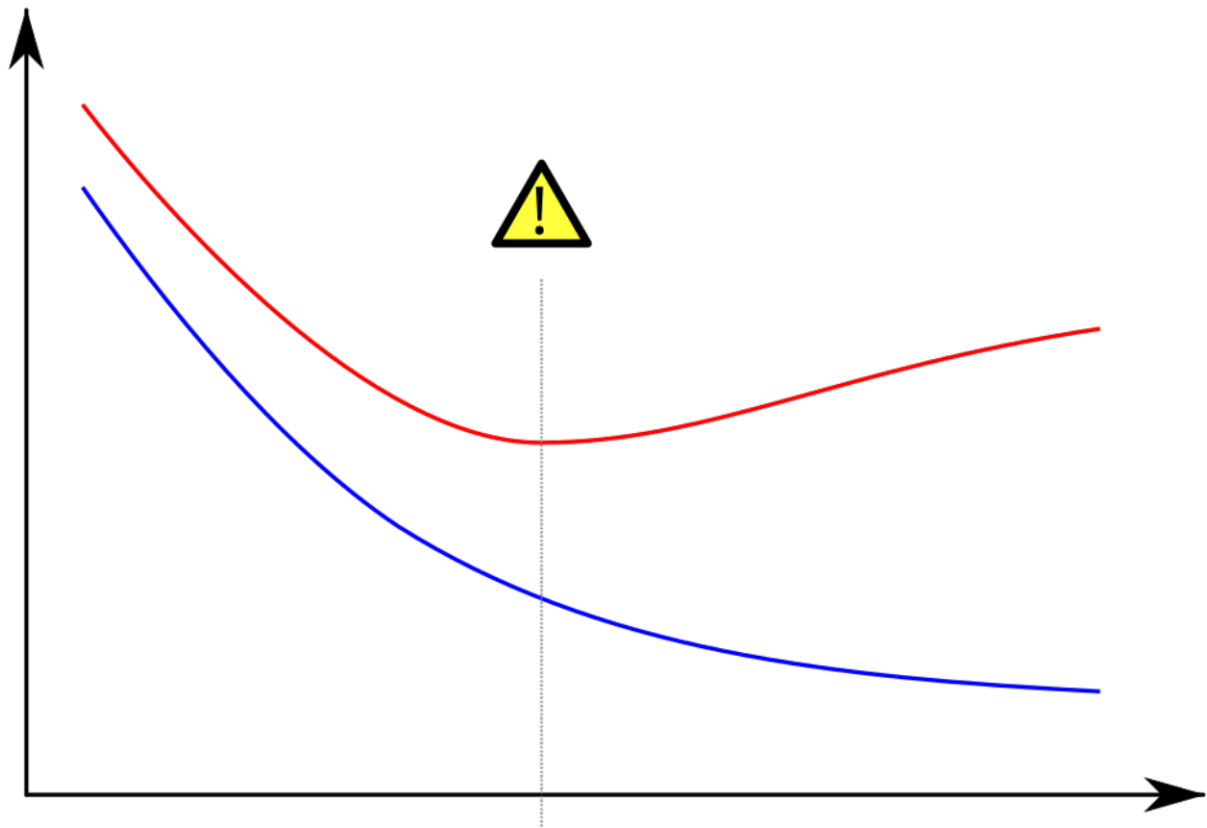
- ☐ SARSA assumes we are following a greedy policy when calculating Q-value up  as
- ☐ SARSA uses a learning rate in its Q-value calculation
- ☐ SARSA is an example of on-policy learning
- ☐ SARSA iteratively updates its estimates of Q

Maximum marks: 2

13 You have a dataset with 1000 instances in total, and you decide to use **five-fold cross-validation** to split it into training and validation sets, so that there will be a total of **five different data splits**. (You can assume that the final test set comes from a different source and is not affected by cross-validation.) Which **two** of the following statements are **true**?
Select exactly two alternatives:

- ☐ The training set of split 2 and the validation set of split 2 have 200 common instances.
- ☐ The training set of split 1 and the validation set of split 2 have 100 common instances.
- ☐ The training set of split 1 and the validation set of split 2 have 0 common instances.
- ☐ The training set of data split 1 contains 900 instances.
- ☐ The training set of split 1 and the validation set of split 2 have 200 common instances. ✓
- ☐ The training set of data split 2 contains 800 instances. ✓

Maximum marks: 2



Which statement is **true** about the image above?

Select one alternative:

- ☐ The image shows an example of underfitting, where the blue line represents the validation loss and the red line the training loss.
- ☐ The image shows an example of overfitting, where the blue line represents the training loss and the red line the validation loss. ✓
- ☐ The image shows an example of underfitting, where the blue line represents the training loss and the red line the validation loss.
- ☐ The image shows an example of overfitting, where the blue line represents the validation loss and the red line the training loss.

Maximum marks: 2

15 What is the value of the **L2 norm/distance** of the weight vector $w = [1, 0.5, 3, 0.4]$?
Note: Some variants of the L2 norm/distance apply a square root, others do not. In this question, we want you to use the variant **without the square root**.
Select one alternative:

☐ 4.9

☐ 10.41



☐ 9.8

☐ 9.6

Maximum marks: 2

16 Which **two** statements are **true**?
Select exactly two alternatives:

☐ In a kNN model, the smaller the value of k, the smaller the bias.



☐ Adding regularization to a logistic regression model increases the bias.



☐ Adding regularization to a logistic regression model increases the variance.

☐ In a kNN model, the smaller the value of k, the smaller the variance.

Maximum marks: 2

17 You have a kNN-based classifier, a logistic regression classifier and a feed-forward neural network with 2 hidden layers. If you want to combine them into one **voting classifier**, you can use:

Select one alternative:

- ☐ soft vote
- ☐ both soft vote and hard vote
- ☐ neither soft vote nor hard vote
- ☐ hard vote



Maximum marks: 2

18 Which statement is **true** about **deep neural networks**?

Select one alternative:

- ☐ ReLU is commonly used as the activation function in a hidden layer.
- ☐ Softmax is commonly used as the activation function in a hidden layer.
- ☐ There is usually no activation function in a hidden layer.
- ☐ The dot product is commonly used as the activation function in a hidden layer.



Maximum marks: 2

19 Which **two** statements are **true**?

Select exactly two alternatives:

- ☐ Recurrent neural networks (RNNs) only support text data on the input side but cannot generate text.
- ☐ RNNs are difficult to train because they suffer from vanishing gradients. ✓
- ☐ Convolutional neural networks (CNNs) are particularly well suited for processing images. ✓
- ☐ LSTMs are a type of convolutional neural network.

Maximum marks: 2

20 Which **linkage method** relies on within-cluster variance when deciding which clusters to merge?

Select one alternative:

- ☐ Average linkage
- ☐ Ward ✓
- ☐ K-means
- ☐ Single linkage

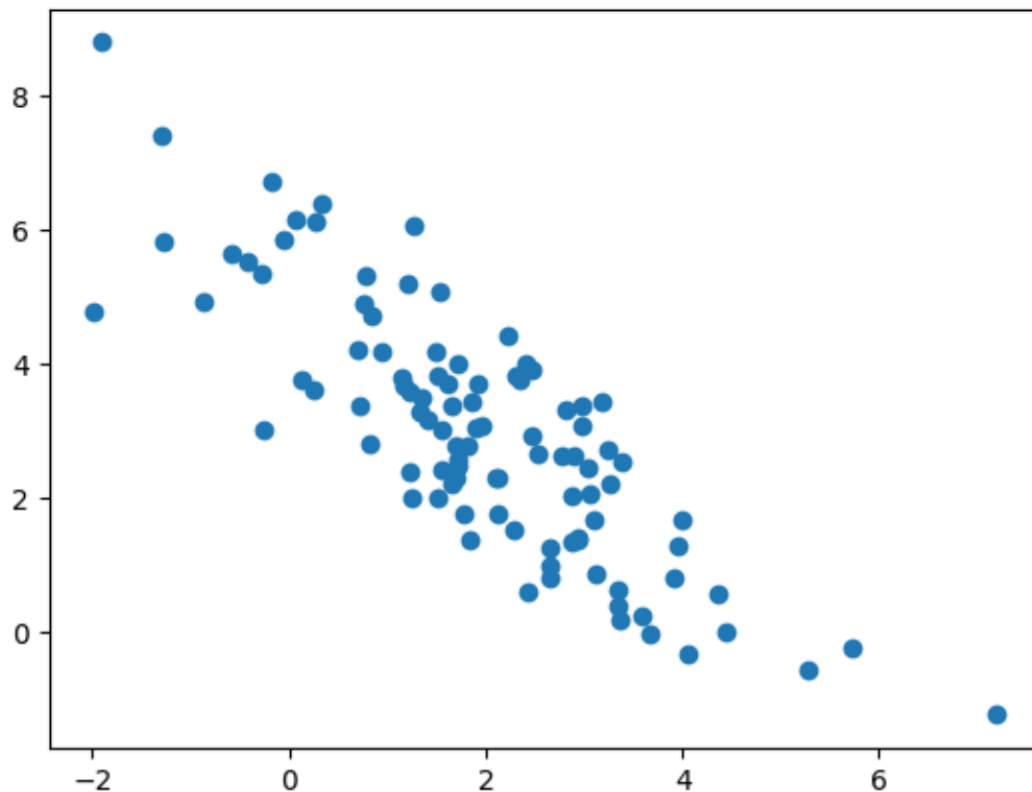
Maximum marks: 2

21 Which **two** statements are **true** about **autoencoders**?

Select exactly two alternatives:

- ☐ Variational autoencoders can be used to generate synthetic data similar to the input. ✓ t.
- ☐ Autoencoders are a type of autoregressive model.
- ☐ Autoencoders can be used to produce compressed representations of the input. ✓ a.
- ☐ Denoising autoencoders use the same input and output data.

Maximum marks: 2



You have a data set with sample points represented as two-dimensional vectors, as illustrated in the image above, where the first dimension is referred as x and the second as y . You apply **principal component analysis (PCA)** to this data set. Which statement is **true**?

Select one alternative:

- ☐ Rotating all sample points by 90° in clockwise direction changes the result of the PCA. ✓
- ☐ Running the PCA algorithm with a different random seed changes the result of the PCA.
- ☐ Moving all sample points by -1 unit on the x axis and +1 unit on the y axis changes the result of the PCA.
- ☐ Moving all sample points by +1 unit on the x axis changes the result of the PCA.

Maximum marks: 2

23 Which **two** statements are **true** about **symbolic and rule-based AI**?

Select exactly two alternatives:

- ☐ Is based on the computational theory of mind. ✓
- ☐ Can pass the Turing test more easily than other AI approaches.
- ☐ Can only be successfully applied to limited domains. ✓
- ☐ Focuses on induction rather than deduction.

Maximum marks: 2

24 Which statement describes the concept of **connectionism** best?

Select one alternative:

- ☐ Connectionism states that neural networks should only have fully connected hidden layers.
- ☐ Networks without hidden layers can only solve linearly separable problems, therefore connectionism is not seen as a promising approach to artificial intelligence.
- ☐ Each neuron in the human brain has a specific function, therefore each neuron in an artificial neural network should also have a specific function.
- ☐ The human brain consists of a huge neural network, therefore artificial intelligence is best modelled with an artificial neural network. ✓

Maximum marks: 2

25 When evaluated on a test set, your classifier yields 10 true positives (TP), 3 false positives (FP) and 4 false negatives (FN). What is the F1 score of your classifier?

Select one alternative:

☐ 5.41

☐ 0.37

☐ 0.74



☐ 0.77

Maximum marks: 2

26 What is the simplest and fastest to compute of the following non-linear activation functions?

Select one alternative:

☐ Hyperbolic tangent

☐ Sigmoid

☐ Softmax

☐ Rectified linear unit (ReLU)



Maximum marks: 2

27 What is true about standard scaling of the features in the context of supervised machine learning?

Select one alternative:

☐ Scaling usually makes the classifier more computationally efficient.



☐ Scaling always improves the classifier performance

☐ You cannot scale the data if you don't know their true labels

☐ Scaling makes data more interpretable

Maximum marks: 2

28 For a classification task, you are using a standard perceptron with step activation function. Its weights are [4, 0.5, 10]. What value will this perceptron produce at the output, if the input is [-1, 4, 0]?

Select one alternative:

☐ 1

☐ 8

☐ 0



☐ -2

Maximum marks: 2

29 What does it mean for a classifier to be “linear”?

Select one alternative:

☐ It trains in linear time to the amount of training data

☐ Its inference (prediction) time is linear to the amount of test data

☐ Its decision boundary is always linear.



☐ Its training loss plot looks like a straight line.

Maximum marks: 2

30 You are testing a linear regression system on two test data points. The correct ("gold") values for them are 4 and -1. The system's predictions are 3 and -5 correspondingly. What is the mean square error (MSE) of this system?

Select one alternative:

☐ 1.0

☐ 8.5



☐ 0.0

☐ 5.5

Maximum marks: 2

31 To solve a 3-class classification problem (classes 1, 2 and 3), you have trained a multinomial logistic regression classifier. For a specific test data point belonging to class 2, your classifier produces the following output: [0.035, 0.260, 0.705]. What is the cross-entropy loss on this data point (the values in the answer options are approximate)?

Select one alternative:

☐ 1/3

☐ 0.35

☐ -1.35

☐ 1.35



Maximum marks: 2

32 This is an update rule for weights of a classifier:

$$w_{i,j} = w_{i,j} - \eta(y_j - t_j)x_i$$

In this rule, what does η ("eta") denote?

Select one alternative:

- ☐ gradient
- ☐ target label
- ☐ learning rate
- ☐ loss



Maximum marks: 2

33 What is the derivative of the function $f(x) = c$ (where c is a scalar number)?

Select one alternative:

- ☐ 0
- ☐ x^2
- ☐ x
- ☐ 1



Maximum marks: 2

34 What is **true** about the backpropagation algorithm?

Select one alternative:

- ☒ It updates weights of a model to decrease its loss. ✔
- ☐ Backpropagation is computationally efficient, because it can skip layers of an artificial neural networks.
- ☐ It is a part of single layer perceptron training pipeline.
- ☐ It is used when testing artificial neural networks on unseen data.

Maximum marks: 2

35 You are given a feed-forward neural network (FFNN). It performs multinomial classification of the input data instances into three classes. No bias term is used. The weight values of the last layer are given below. What class this network will predict if the input to the last layer is $h=[2, -1, 4, 0]$?

2 1 1

2 2 1

1 2 2

1 1 2

Select one alternative:

- ☒ Class 3 ✔
- ☐ Class 1
- ☐ Equal probability for all classes
- ☐ Class 2

Maximum marks: 2

36 *Your responses in text fields are saved automatically*

- a)** Explain what do the terms “feature” and “label” mean in the context of supervised machine learning.
- b)** What feature types do you know?
- c)** How they are related to the notion of “representation”?
- d)** Come up with an example of a classification problem, pointing out what exactly serve as features and labels in this specific problem.

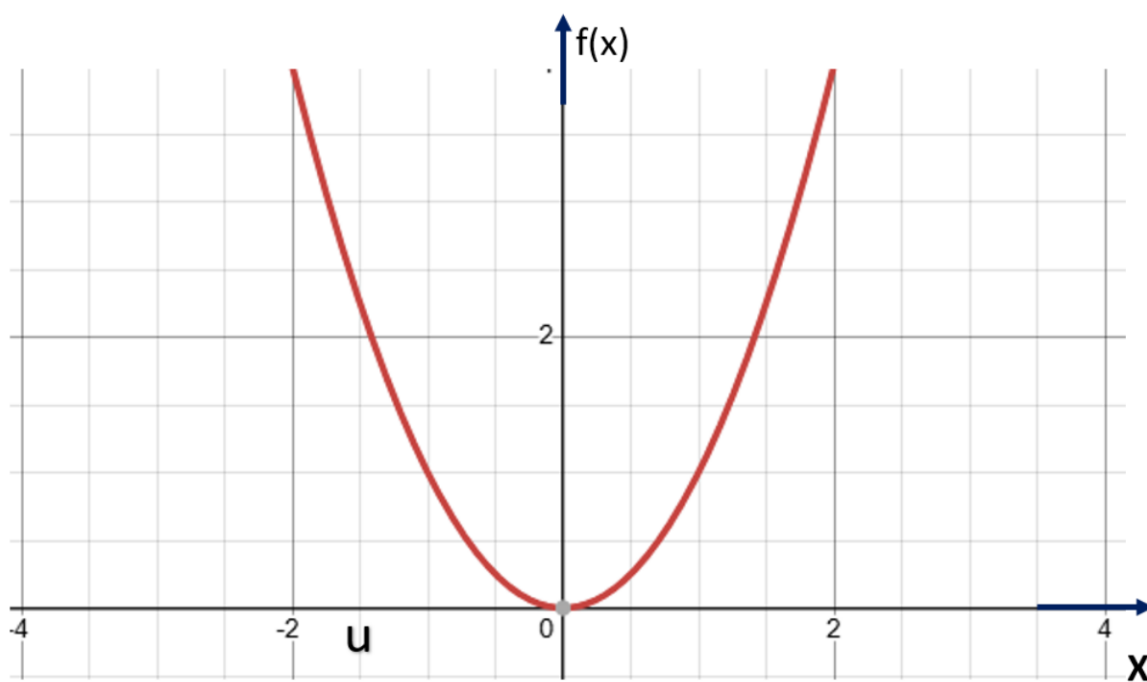
Add your free text answer

Maximum marks: 6

37 The figure below shows a 1-dimensional search landscape for a *minimization* problem. We are optimizing a variable x , and the quality of our solution varies as shown depending on the value of x . We start searching for a solution at the value $x=u$.

Different search methods have different ways of finding the next solution to evaluate. Let's call the next evaluated point in the search landscape v . Describe in words **how v will be selected**, given the search methods

- Gradient descent. Mention how v depends on the descent rate/ step size (called α in the course textbook by Marsland) (3.5p)
- Local search / hillclimbing (or, more accurately hill descending in our case). State any assumptions you make around how neighbor solutions are explored (what Marsland calls the move-set). (3.5p)



Fill in your answer here

Words: 0

Maximum marks: 7

38 *Your responses in text fields are saved automatically*

When training a model using gradient descent, one can use several different approaches: stochastic gradient descent (SGD), batch training and mini-batch training (among others). Describe these varieties, their advantages and disadvantages. Can some of them be considered a subset of others?

Add your free text answer

Maximum marks: 6

- 39 a) What is the role of the discount factor in Reinforcement Learning? (2.5p)
b) Suggest a Reinforcement Learning scenario where it would be wise to have a relatively low discount factor. Briefly state the reasoning behind your answer. (2.5p)

Fill in your answer here

Maximum marks: 5

- 40** *Your responses in text fields are saved automatically*

In the history of AI, it has been common practice to evaluate new AI approaches and models on their performance in playing board games such as Chess or Go. Provide two arguments in favour of, and two arguments against this evaluation method.

Fill in your answer here

Maximum marks: 6